

Alphabet CEO Sundar Pichai speaking at the Google I/O conference on May 10th, 2023. COURTESY OF GOOGLE

**THIS WEEK, AT** its annual I/O developer conference in Mountain View, Google showcased a head-spinning number of projects and products powered by or enhanced by AI. They included a new-and-improved version of its chatbot Bard, tools to help you write emails and documents or manipulate images, devices with AI baked in, and a chatbot-like experimental version of Google search. For a full recap of the event, complete with insightful and witty commentary from my WIRED colleagues, check out our Google I/O liveblog.



What's New With WIRED



## The Boring Future of Generative AI

00:00 / 05:35



Google's big pivot is, of course, largely fueled not by algorithms but by generative AI FOMO.

---

**Sign Up Today**

This is an edition of WIRED's Fast Forward newsletter, a weekly dispatch from the future by Will Knight, exploring AI advances and other technology set to change our lives.

---

The appearance last November of ChatGPT—the remarkably clever but still rather flawed chatbot from OpenAI—combined with Microsoft adding the technology to its search engine Bing a few months later, triggered something of a panic at Google. ChatGPT proved wildly popular with users, demonstrating new ways to serve up information that threatened Google’s vice grip on the search business and its reputation as the leader in AI.

The capabilities of ChatGPT and AI language algorithms like those powering it are so striking that some experts, including Geoffrey Hinton, a pioneering researcher who recently left Google, have felt compelled to warn that we might be building systems

that we will someday struggle to control. OpenAI's chatbot is often astonishingly good at generating coherent text on a given subject, summarizing information from the web, and even answering extremely tricky questions that require expert knowledge.

And yet, unfettered AI language models are also silver-tongued agents of chaos. They will gladly fabricate facts, express unpleasant biases, and say unpleasant or disturbing things with the right prompting. Microsoft was forced to limit the capabilities of Bing chat shortly after launch to avoid such embarrassing misbehavior, in part because its bot divulged its secret codename—Sydney—and accused a *New York Times* columnist of not loving his spouse.

Google worked hard to tone down the chaotic streak of text-generation technology as it prepared the experimental search feature announced yesterday that responds to search queries with chat-style answers synthesizing information from across the web.

Google's smarter version of search is impressively narrow-minded, refusing to use the first person or talk about its thoughts or feelings. It completely avoids topics that might be considered risky, refusing to dispense medical advice or offer answers on potentially controversial topics such as US politics.

Google deserves recognition for reining in generative chatbots' wild side like that. But in my tests, the new search interface felt incredibly tame compared to ChatGPT or Google's own chatbot Bard.

As the company moves the technology into more of its products, perhaps the generative AI revolution will turn out to be a lot less fun than you might expect from the early shock and awe of ChatGPT, a chatbot that has an edgy charm. Gone are the wild ravings and imaginings of powerful AI bots. In their place are new ways to populate spreadsheets, compose email pleasantries, and find products to buy.

Even if "AI doomers" warning about errant AI prove overblown, it will be interesting to watch how companies like Google and OpenAI balance the development of more powerful generative language models with the need to have them behave.

---

### See What's Next in Tech With the Fast Forward Newsletter

A weekly dispatch from the future by Will Knight, exploring AI advances and other technology set to change our lives. Delivered every Thursday.

---

By signing up you agree to our [User Agreement](#) (including the [class action waiver and arbitration provisions](#)), our [Privacy Policy & Cookie Statement](#) and to receive marketing and account-related emails from WIRED. You can unsubscribe at any time.



Google has invested huge sums and major resources in AI over recent years, with CEO Sundar [Pichai often pitching the company as “AI first”](#), and the company is desperate to show it can advance the technology more quickly than OpenAI. One high-level message from Google's stream of AI announcements was that the company is not going to hold back anymore, as it did the LaMDA chatbot that was [announced long before ChatGPT](#) appeared but not made public.

In March, some big names in AI research signed an open letter calling for a six-month pause on creating machine learning systems more powerful than GPT-4, which powers ChatGPT. Pichai was not a signatory and said in his keynote speech yesterday that the company is currently training a new, more powerful language model called Gemini.

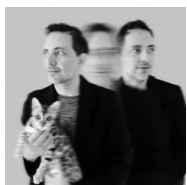
A source at Google tells me this new system will incorporate a range of recent advances from different large language models and may eclipse GPT-4. But don't expect to get to experience the full power or charisma Gemini can offer. If Google applies the same chaos-taming methods seen in its chat-like search experiment, it may just seem like another surprisingly clever autocomplete.

---

## Get More From WIRED

-  Understand AI advances with our [Fast Forward newsletter](#)
-  Our new [podcast](#) wants you to [Have a Nice Future](#)
- It's the Age of Ozempic. Do we still [need Weight Watchers?](#)
- [Temu is losing millions](#) of dollars to send you cheap socks
- Sex workers are still hot for [Twitter Spaces](#)
- The security hole at the heart of [ChatGPT and Bing](#)
- [Remembering GitHub's office](#), a monument to tech culture

- 📷 Snap into spring with the Gear team's picks for the best camera bags, fun instant cameras, and mirrorless cameras



Will Knight is a senior writer for WIRED, covering artificial intelligence. He was previously a senior editor at *MIT Technology Review*, where he wrote about fundamental advances in AI and China's AI boom. Before that, he was an editor and writer at *New Scientist*. He studied anthropology and journalism in... [Read more](#)

SENIOR WRITER



---

TOPICS   FAST FORWARD   ARTIFICIAL INTELLIGENCE   SEARCH   GOOGLE   MACHINE LEARNING   CHATGPT

---

MORE FROM WIRED

---

**Google Just Added Generative AI to Search**

ARTIFICIAL INTELLIGENCE / TECH / OP-ED

## OpenAI isn't doing enough to make ChatGPT's limitations clear /

Users deserve blame for not heeding warnings, but OpenAI should be doing more to make it clear that ChatGPT can't reliably distinguish fact from fiction.

---

By [James Vincent](#), a senior reporter who has covered AI, robotics, and more for eight years at The Verge.

May 30, 2023, 11:22 AM EDT

☐ [28 Comments](#) / [28 New](#)



Illustration: The Verge

“May occasionally generate incorrect information.”

This is the warning OpenAI pins to the homepage of its AI chatbot ChatGPT — one point among nine that detail the system’s capabilities and limitations.

“May *occasionally* generate incorrect information.”

It’s a warning you could tack on to just about any information source, from Wikipedia to Google to the front page of *The New York Times*, and it would be more or less correct.

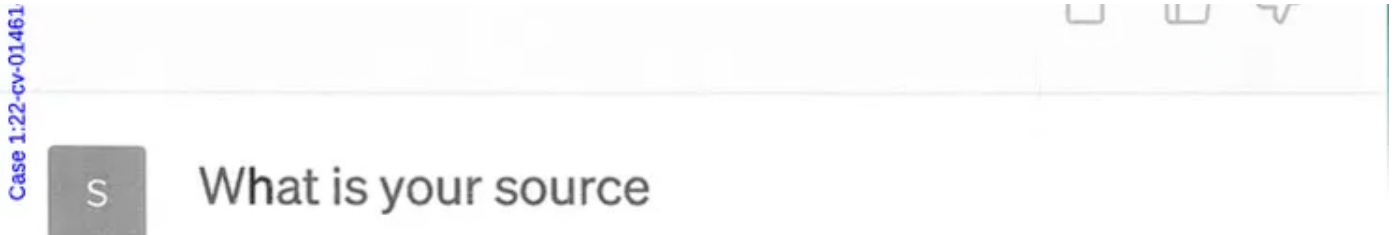
“May occasionally generate *incorrect* information.”

Because when it comes to preparing people to use technology as powerful, as hyped, and as misunderstood as ChatGPT, it’s clear OpenAI isn’t doing enough.

The misunderstood nature of ChatGPT was made clear for the umpteenth time this weekend when news broke that US lawyer Steven A. Schwartz had turned to the chatbot to find supporting cases in a lawsuit he was pursuing against Colombian airline Avianca. The problem, of course, was that none of the cases ChatGPT suggested exist.

Schwartz claims he was “unaware of the possibility that [ChatGPT’s] content could be false,” though transcripts of his conversation with the bot show he was suspicious enough to check his research. Unfortunately, he did so by asking ChatGPT, and again, the system misled him, reassuring him that its fictitious case history was legitimate:





*"Don't believe his lies."* Image: SDNY

Schwartz deserves plenty of blame in this scenario, but the frequency with which cases like this are occurring — when users of ChatGPT treat the system as a reliable source of information — suggests there also needs to be a wider reckoning.

Over the past few months, there have been numerous reports of people being fooled by ChatGPT's falsehoods. Most cases are trivial and have had little or no negative impact. Usually, the system makes up a news story or an academic paper or a book, then someone tries to find this source and either wastes their time or looks like a fool (or both). But it's easy to see how ChatGPT's misinformation could lead to more serious consequences.

In May, for example, one Texas A&M professor used the chatbot to check whether or not students' essays had been written with the help of AI. Ever obliging, ChatGPT said, yes, *all* the students' essays were AI-generated, even though it has no reliable capability to make this assessment. The professor threatened to flunk the class and



withhold their diplomas until his error was pointed out. Then, in April, a law professor recounted how the system generated false news stories accusing him of sexual misconduct. He only found out when a colleague, who was doing research, alerted him to the fact. “It was quite chilling,” the professor told *The Washington Post*. “An allegation of this kind is incredibly harmful.”

I don’t think cases like these invalidate the potential of ChatGPT and other chatbots. In the right scenario and with the right safeguards, it’s clear these tools can be fantastically useful. I also think this potential includes tasks like retrieving information. There’s all sorts of interesting research being done that shows how these systems can and will be made more factually grounded in the future. The point is, right now, it’s not enough.

This is partly the fault of the media. Lots of reporting on ChatGPT and similar bots portray these systems as human-like intelligences with emotions and desires. Often, journalists fail to emphasize the unreliability of these systems — to make clear the contingent nature of the information they offer.

## **People use ChatGPT as a search engine. OpenAI needs to recognize that and warn them in advance**

But, as I hope the beginning of this piece made clear, OpenAI could certainly help matters, too. Although chatbots are being presented as a new type of technology, it’s clear people use them as search engines. (And many are explicitly launched as search engines, so of course they get confused.) This isn’t surprising: a generation of internet users have been trained to type questions into a box and receive answers. But while sources like Google and DuckDuckGo provide links that invite scrutiny, chatbots muddle their information in regenerated text and speak in the chipper tone of an all-knowing digital assistant. A sentence or two as a disclaimer is not enough to override this sort of priming.

Interestingly, I find that Bing’s chatbot (which is powered by the same tech as ChatGPT) does slightly better on these sorts of fact-finding tasks; mostly, it tends to search the web in responses to factual queries and supplies users with links as

sources. ChatGPT *can* search the web, but only if you're paying for the Plus version and using the beta plug-ins. Its self-contained nature makes it more likely to mislead.

Interventions don't need to be complex, but they need to be *there*. Why, for example, can ChatGPT simply not recognize when it's being asked to generate factual citations and caution the user to "check my sources"? Why can't it respond to someone asking "is this text AI-generated?" with a clear "I'm sorry, I'm not capable of making that judgment"? (We reached out to OpenAI for comment and will update this story if we hear back.)

OpenAI has definitely improved in this area. Since ChatGPT's launch, it has, in my experience, become much more up-front about its limitations, often prefacing answers with that AI shibboleth: "As an AI language model..." But it's also inconsistent. This morning, when I asked the bot "can you detect AI-generated text?" it cautioned that it was "not foolproof," but then when I fed it a chunk of this story and asked the same question, it simply replied: "Yes, this text was AI-generated." Next, I asked it to give me a list of book recommendations on the topic of measurement (something I know a little about). "Certainly!" it said before offering 10 suggestions. It was a good list, hitting many of the classics, but two titles were just entirely made up, and if I hadn't known to check, I wouldn't have noticed. Try similar tests yourself and you'll quickly find errors.

With this sort of performance, a disclaimer like "May occasionally generate incorrect information" just doesn't seem accurate.

28 COMMENTS (28 NEW)

FEATURED VIDEOS FROM THE VERGE

## TikTok Hearing

